Conditioned Task-set Competition: Neural Mechanisms of Emotional Interference in Depression

Cogn Affect Behav Neurosci
DOI 10.3758/s13415-016-0478-4

Conditioned task-set competition: Neural mechanisms of emotional interference in depression

Aleks Stolicyn[1] · **J. Douglas Steele**[2] · **Peggy Seriès**[1]

Abstract Depression has been associated with increased response times at the incongruent-, neutral-, and negative-word trials of the classical and emotional Stroop tasks (Epp et al., *Clinical Psychology Review, 32,* 316–328, 2012). Response-time slowdown effects at incongruent- and negative-word trials of the Stroop tasks were reported to correlate with depressive severity, indicating strong relevance of the effects to the symptomatology. This study proposes a novel integrative computational model of neural mechanisms of both the classical and emotional Stroop effects, drawing on the previous prominent theoretical explanations of performance at the classical Stroop task (Cohen, Dunbar, & McClelland, *Psychological Review, 97,* 332–361, 1990; Herd, Banich, & O'Reilly, *Journal of Cognitive Neuroscience, 18,* 22–32, 2006), and in addition suggesting that negative emotional words represent conditioned stimuli for future negative outcomes. The model is shown to explain the classical Stroop effect and the slow (between-trial) emotional Stroop effect with biologically plausible mechanisms, providing an advantage over the previous theoretical accounts (Matthews & Harley, *Cognition & Emotion, 10,* 561–600, 1996; Wyble, Sharma, & Bowman, *Cognition & Emotion, 22,* 1019–1051, 2008). Simulation results suggested a candidate mechanism responsible for the pattern of depressive performance at the classical and the emotional Stroop tasks. Hyperactivity of the amygdala, together with increased inhibitory influence of the amygdala over dopaminergic neurotransmission, could be at the origin of the performance deficits.

Keywords Depression · Amygdala · Computational model · Emotion · Dopamine · Neural network

✉ Peggy Seriès
pseries@inf.ed.ac.uk

[1] Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, UK

[2] School of Medicine (Neuroscience), University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK

Introduction and background

Classical and emotional Stroop tasks are experimental paradigms which probe cognitive control in the face of conflicting information and emotional distraction, respectively. Depression is associated with performance deficits at both tasks (Epp et al., 2012). In this study, we first set out to investigate the neural mechanisms underpinning the classical and emotional Stroop effects, from a theoretical perspective. We then proceed to suggest deficits in these mechanisms which might be characteristic of depression.

Classical Stroop effect and neural mechanisms

The classical Stroop task requires participants to name the ink colour of a word, where the verbal meaning of the word itself is either colour-congruent (e.g. *Red* printed in red ink), incongruent (e.g. *Green* printed in red ink), or not relevant to the response (neutral; e.g. *Glass* printed in red ink). The hallmark Stroop effect manifests itself as the delay in response when colour-naming incongruent combinations, compared to neutral combinations. Quickest responses are observed to congruent stimuli. Slower responses in the incongruent condition indicate an added cognitive load. As a result, the task has been used for investigating mechanisms of cognitive control (Cohen, Dunbar, & McClelland, 1990; MacLeod, 1991).

From a mechanistic perspective, arguably the most influential model of the classical Stroop effect has been proposed

Published online: 09 December 2016

by Cohen et al. (1990) within the connectionist parallel distributed processing (PDP) framework. The model posits two parallel competing neural pathways in the brain—one dedicated to processing colour and another to processing words. Because word-reading behaviour is more frequently practiced than colour naming, the word-reading pathway is theorized as being overtrained compared to the colour-naming pathway. When the network performs the colour-naming task, the word-reading pathway provides strong competition due to its automatic habitual nature. To overcome this competition, the model proposes task-specific facilitation nodes, corresponding to neural representations of executive control. During colour naming, the active colour-naming executive task set facilitates the colour-naming processing pathway, which is then able to overcome the word-reading pathway competition and activate correct responses. The executive facilitation, however, is not strong enough to compensate completely for the overtraining effects within the word-reading pathway. This results in slower, although correct, performance at colour naming, compared to word reading (with the strongest effect at the incongruent trials)—as has been reported experimentally. The Cohen et al. (1990) model accounts successfully for the classical Stroop effect.

A notable extension of the PDP Stroop model has been proposed in Herd, Banich, and O'Reilly (2006), termed the top-down excitatory bias (TEB) model of the Stroop effect. The original account by Cohen et al. (1990) suggested long-range inhibitory connections between the executive and the processing areas. These connections, however, appeared biologically implausible. Banich et al. (2000) and Banich et al. (2001) have also revealed some rather counterintuitive neuroimaging results related to the classical Stroop task. Specifically, BOLD activations in verbal processing areas appeared higher in incongruent trials compared to neutral trials. The original PDP model could not account for these findings. Herd et al. (2006) have improved the model to exclude long-range inhibitory connections and include representations of categories (alongside task units), responsible for facilitation of colour-related information in all processing pathways. With the new colour category representation, the model could account for the pattern of verbal area activations reported in Banich et al. (2000, 2001).

Common neurobiological interpretations of the PDP and TEB models posit that the dorsolateral prefrontal cortex (DLPFC) is responsible for maintaining representations of the task set. This is supported by the experimental results (Nee, Wager, & Jonides, 2007). Parietal-cortical verbal and colour processing areas are proposed to correspond to the colour and word processing pathways in the models (Cohen et al., 1996; Herd et al., 2006).

Emotional Stroop effect and neural mechanisms

The emotional Stroop task, in contrast to the classical Stroop task, uses affective (positive, negative, or neutral) rather than colour words, with the similar task for the participants to name the colour of the ink. The main finding is that negative words, compared to positive and neutral words, cause interference with task performance, measured with increased response times (e.g. Algom, Chajut, & Lev, 2004; Frings, Englert, Wentura, & Bermeitinger, 2010; McKenna & Sharma, 2004; see review in Phaf & Kan, 2007). An important difference between the two tasks should be noted: whereas ink colour and word meaning are designed to induce *response conflict* in the classical task, no conflict is present in the emotional task. Instead, the response delay is considered to arise because of the emotional relevance of the words. Whereas the classical Stroop effect appears to be strongly manifested immediately (i.e. in the trials with ink-incongruent colour words; MacLeod, 1991), the emotional Stroop effect appears to be expressed more in a carry-over fashion (i.e. in the trials *immediately following* those with negative words; e.g. Algom et al., 2004; McKenna & Sharma, 2004). A meta-analysis has shown that the effect is much more pronounced when negative-word trials are presented in blocks rather than intermixed with neutral words (Phaf & Kan, 2007). This suggests that negative words induce a between-trial slowdown. The task has been useful for investigating the neural basis of control over emotional interference (e.g. Compton et al., 2003), as well as disturbances in anxiety and depression (e.g. Gotlib & McCann, 1984; Williams, Mathews, & MacLeod, 1996; Mitterschiffthaler et al., 2008).

Compared to the classical Stroop effect, relatively few studies have looked at the neural basis of the emotional Stroop effect. Crucially, activation of the amygdala has been reported in generation of the effect (Isenberg et al., 1999). Lack of behavioural slowdowns at negative-word trials, on the other hand, was accompanied by activation of the DLPFC, and deactivation of the amygdala (Compton et al., 2003). Activated amygdala has also been highlighted at other tasks during emotional word processing (Hamann & Mao, 2002; Naccache et al., 2005; Straube, Sauer, & Miltner, 2011), and during emotional distraction at executive and attention tasks (see review in Iordan, Dolcos, & Dolcos, 2013). In the latter case, amygdala appeared activated together with ventral prefrontal cortex, and accompanied by deactivation of the executive control areas, including the DLFPC. Apart from the amygdala, several studies reported activations of the rostral anterior cingluate cortex (rACC; Mohanty et al., 2007; Whalen et al., 1998).

Compared to the classical Stroop task, only a single computational modelling study to date accounts for the mechanistic basis of the *slow* emotional Stroop effect. Wyble et al. (2008) expand on the earlier conflict resolution account by

Botvinick and colleagues (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Yeung, Botvinick, & Cohen, 2004). They suggest that automatic processing of negative emotional words results in decreased deployment of attentional/cognitive control necessary for the ongoing task. This should release resources for preparing to locate and process a potential threat and lead to a delay in colour naming after negative words, corresponding to the slow emotional Stroop effect. As in the previous modelling studies, the authors suggest that the cognitive control node might correspond to the dorsal anterior cingulate cortex (dACC). The authors further suggest that the rACC could be recruited when processing negative or threatening material, with a specific role to inhibit the dACC and reduce task-related attention. The model is consistent with evidence of involvement of the rACC in generating the emotional Stroop effect (Bush, Luu, & Posner, 2000; Whalen et al., 1998). It also accounts well for the slow emotional Stroop effect (Algom et al., 2004; McKenna & Sharma, 2004). The neurobiological interpretation of the model, however, remains controversial. Mohanty et al. (2007), for example, reported correlating activations in the rACC and the dACC during performance at the emotional Stroop task, which contradicts the competition hypothesis. Etkin, Egner, Peraza, Kandel, and Hirsch (2006) and Egner, Etkin, Gale, and Hirsch (2008) reported an inhibitory effect of the rACC activation over the subsequent amygdala activity, in a face–word version of the task. This suggests that the rACC is involved in diminishing bottom-up propagation of emotional information—a theory opposite to Wyble et al. (2008).

Depressive classical and emotional Stroop effects

Depression is a prevalent psychiatric disorder characterized by a range of affective and cognitive symptoms, including persistent sad mood or depressive ruminative thought, diminished ability to concentrate, and diminished ability to make decisions (American Psychiatric Association, 1994). A recent meta-analysis has indicated a robust effect of depression on performance in the classical and the emotional Stroop tasks (Epp et al., 2012). The authors reviewed 47 studies and reported that depression is robustly associated with higher colour-naming response times in all except congruent conditions, with negatively valenced words being associated with the strongest effect compared to neutral and positive words. Significantly stronger interference (depression compared to controls) was also found in the incongruent condition of the classical Stroop task. Results are consistent with a previous meta-analysis of attentional bias in depression (Peckham, McHugh, & Otto, 2010). Most notably, Epp and colleagues have discovered a correlation between depressive symptom severity (Hamilton Rating Scale for Depression scores) and depression effect sizes at the negative and incongruent Stroop task conditions. This suggests that changes in the Stroop performance in depression, particularly in association with negatively valenced stimuli, could be relevant to the symptomatology of the disorder.

Few studies have investigated the neural basis of altered performance at the classical Stroop task in depression. One fMRI study indicated increased activations in the rACC and the DLPFC in unmedicated depressed patients (Wagner et al., 2006). In contrast, decreased activations in many brain regions—including the middle frontal gyrus and the posterior cortex—were reported at incongruent condition in another study (Kikuchi et al., 2012). In both cases, however, no behavioural differences between depressed and control participants were observed. When increased response times and error rates *were* observed, they were reported to correlate with decreased activations in small regions of the DLPFC and the dACC (Holmes & Pizzagalli, 2008). Overall, these results indicate that lower activation in the prefrontal regions (including the DLPFC) might contribute to slower performance in the classical Stroop task, while overactivation of these region might represent compensatory activity.

With regard to the neural correlates of the *emotional* Stroop task performance in depression, only a single fMRI study appears to have been conducted. Results indicate hyperactivity in the rACC and in the precuneus, with the rACC hyperactivation positively correlating with the negative-word trial response latencies (Mitterschiffthaler et al., 2008). The authors did not find stronger activation of the amygdala (reported to be hyperactive in depression; Drevets, 2003; Hamilton et al., 2012; Whalen, Shin, Somerville, McLean, & Kim, 2002). Drawing on Etkin et al. (2006), Mitterschiffthaler et al. (2008) suggested that the rACC could act as a compensatory mechanism—to inhibit lower-level emotional processing in the amygdala. If this is the case, the hyperactive rACC would not contribute directly to the behavioural effects of depression at the emotional Stroop task.

Depression neurobiology

General neurobiology of depression has been the subject of several important theoretical reviews over the last decades. Mayberg (1997) proposed an influential depression model emphasizing increases in ventral limbic area activations (amygdala, hippocampus, hypothalamus), alongside an activity decrease in dorsal neocortical (dACC and DLPFC) areas. These neural alterations arguably correspond to increased negative conditioned and affective processing (mood symptoms) and decreased task-related cognitive control (cognitive symptoms). DeRubeis, Siegle, & Hollon (2008) have suggested an imbalance in the interactions between the amygdala and the prefrontal cortex (PFC) as a hallmark abnormality in depression (with the amygdala exerting a stronger influence over the PFC), which is ameliorated with successful treatment. Disner and colleagues related depressive neural deficits to

components of Beck's influential cognitive model and suggested amygdalar hyperactivity as the crucial contributing factor to negatively biased information processing in attention, memory, and interpretation (Beck, 1967; Disner, Beevers, Haigh, & Beck, 2011). The dACC and the DLPFC are, on the other hand, suggested as hypoactive—exerting reduced modulating control over the limbic affective processing. In line with these suggestions, Roiser and Sahakian (2013) have also proposed a novel cognitive neuropsychological model of depression, stressing deficient cognitive control from the DLPFC and increased negative emotional bias in the amygdala, and other limbic areas, as some of the core depressive neural abnormalities. A general consensus between these theoretical reviews is that limbic affective areas—most notably the amygdala—become metabolically hyperactive in depression, while dorsal cortical areas responsible for higher cognition become hypoactive.

A wealth of evidence indicates deficiency in monoamine neurotransmission in depression, with particular importance of serotonin (Blier & El Mansari, 2013; Mulinari, 2012). Evidence from the last 2 decades, however, also indicates an important role of dopamine in the disorder. Animal models of depression, for example, have been characterized by reduced dopaminergic neurotransmission, particularly in the mesolimbic pathway (Cabib & Puglisi-Allegra, 1996; Gessa, 1996). Pharmacological agents which increase dopamine levels (bupropion and amineptine) are currently used as a second-line treatment for depression with some degree of success (IsHak et al., 2009; Rampello, Nicoletti, & Nicoletti, 2000; Shultz & Malone, 2013). Recent reviews highlight that dopamine transmission alterations are highly relevant to the depressive symptomatology (Dunlop & Nemeroff, 2007; Nestler & Carlezon, 2006; Pizzagalli, 2014). In summary, although serotonin has historically received the most attention in depression, emerging evidence also indicates deficits in dopaminergic neurotransmission.

In this study, we will consider involvement of the amygdala and the dopamine system in the generation of the emotional Stroop effect. We will return to the relevant neurobiological deficits overviewed above when we constrain the possible mechanisms of depression in the Theory and Methods sections.

Modelling aims

Our first aim is to attempt to better explain the neural mechanisms involved in generating the emotional Stroop effects, within an integrative model of both the classical and emotional Stroop tasks. Previous account suggests that the *slow* effect arises because of reduced deployment of cognitive control (Wyble et al., 2008). Its neurobiological interpretation—competition between dACC and rACC—remains controversial (e.g. Etkin et al., 2006; Mohanty et al., 2007). The current modelling study aims to provide a better biologically grounded explanation of the effect drawing on the novel interpretation of emotional words as a case of conditioned stimuli. We suggest that neural mechanisms of conditioned stimuli processing could be responsible for generating the slow emotional Stroop effect.

Our second aim is to investigate mechanisms of the increased response times at both the classical and emotional Stroop tasks in depression. *Increases* in response times at *incongruent* and *negative* trials correlate with symptom severity (Epp et al., 2012). Explanation of neural mechanisms of these deficits can thus indicate core mechanisms of depression (Maia & Frank, 2011).

Theory and modelling methods

Conditioned task-set competition theory

The current study constructs a novel integrative model of the classical and emotional Stroop effects, following the principles outlined in Cohen et al. (1990). We expand the original model with additional biologically-based components to account for the emotional Stroop effect, following interpretation of the emotional words as conditioned stimuli. Briefly, the novel suggestion is that mechanisms of conditioned information appraisal in the brain also generate the emotional Stroop reaction time effects.

Classical (Pavlovian) conditioning refers to the ability to learn associations between neutral stimuli and motivationally salient events—rewards and punishments. The previously neutral stimuli predicting (associated with) rewards or punishments are referred to as conditioned stimuli (CS), while the primary rewards or punishments are referred to as unconditioned stimuli (US). In brief, classical conditioning refers to the ability to learn to evoke behaviours relevant to the US (rewards or punishments) upon a mere presentation of the CS—even when the US is not present. Positive and negative words at the emotional Stroop task could be considered a type of CS because of their primary or secondary associations with motivationally salient concepts or events—rewards or punishments.

Expression of conditioned behaviours is crucially mediated by the amygdala. This is supported by the wealth of rodent studies indicating criticality of the structure for both learning, long-term storage, and expression of conditioned fear (CS–US) associations through the neural mechanism of long-term potentiation (LeDoux, 2003; Maren, 2005; Phelps & LeDoux, 2005). Functional neuroimaging and lesion studies have also supported the role of the amygdala in the expression of fear in humans (Phelps, 2006; Phelps & LeDoux, 2005). Drawing on this existing evidence, in our model the amygdala acts as the primary detector of conditioned material—affective words. It

then sends signals to other brain areas to initiate relevant conditioned behaviours.

Alongside the amygdala, a critical role in conditioned behaviour expression has been reported for regions in the PFC. More specifically, the medial prefrontal cortex (MPFC, encompassing the rACC) is considered to be involved in behavioural expression of conditioned fear, as well as acquisition of fear extinction memories (Courtin, Bienvenu, Einarsson, & Herry, 2013; Marek, Strobel, Bredy, & Sah, 2013; Maroun, 2013). Both the MPFC and the orbitofrontal cortex (OFC, located caudally to the MPFC) have been suggested to evaluate conditioned stimulus signals from the amygdala in order to select and initiate most appropriate instrumental responses (Cardinal, Parkinson, Hall, & Everitt, 2002; Grabenhorst & Rolls, 2011). This is consistent with strong structural interconnections of the regions with the amygdala (Carmichael & Price, 1995; Ray & Zald, 2012). Drawing on this evidence, we suggest that amygdala conditioned stimulus signals (initiated by emotional words) are propagated to the areas in the PFC, where they support representations of behaviours (task sets) relevant for the conditioned material (e.g. escape behaviour in response to a threat word). These conditioned representations are suggested *to draw resources away* from the current ongoing behaviours (task sets).

Complementary evidence also implicates involvement of the dopamine system in conditioned behaviour expression. Dopaminergic burst activity in the ventral tegmental area (VTA) is theorised to represent reward-prediction error signals in the brain (Bromberg-Martin, Matsumoto, & Hikosaka, 2010; Lammel, Lim, & Malenka, 2014). Symmetric to dopaminergic bursts (spikes)—short phasic decreases (dips) in dopamine neuron firing may be characteristic of processing punishments, and conditioned stimuli predictive of punishments (particularly those which are uncontrollable; see reviews in Oleson & Cheer, 2013; Volman et al., 2013). Conditioned dips in firing of the VTA dopamine neurons are likely triggered by inhibitory signals from the immediately posterior rostromedial tegmental area (RMTg; reviewed in Bourdy & Barrot, 2012). The RMTg receives signals from a range of subcortical structures, including, among others, the extended amygdala. A recent optogenetic investigation has indicated that signals from the extended amygdala to the RMTg are sufficient to initiate aversion-related behaviours (Jennings et al., 2013). Drawing on this evidence, we suggest that negative words in the emotional Stroop task are detected as conditioned stimuli by the amygdala, which triggers dopamine *dip* signals in the VTA. Dopamine signals propagate widely in the brain, with the PFC as a major target. Dopamine levels in the PFC have been proposed to promote cognitive stability (stability of the task set), while decreases in dopamine levels should enable flexible shifts of the cognitive tasks (Cools, 2008). Drawing on this theory, we suggest that the amygdala-initiated dopamine dips are propagated to the PFC, where they decrease D1 receptor occupancy (Dreyer, Herrik, Berg, & Hounsgaard, 2010), which subsequently decreases stability of the current task set. This is suggested to enable better processing of the incoming conditioned negative material.

Importantly, dopamine is transmitted to the PFC largely through volume diffusion, rather than directly to synapses (Seamans & Yang, 2004; Lapish, Kroener, Durstewitz, Lavin, & Seamans, 2007). Stable levels of prefrontal dopamine are defined mainly (though not exclusively) by the balance between dopamine release and uptake. Decreased transmission from the VTA to the PFC (dopamine dips) should result in a transient dominance of uptake over release, and thus decreased dopamine occupancy at the receptor sites (Dreyer et al., 2010). Dopamine uptake in the PFC, however, has been shown to be relatively slow, with a time course between one and tens of seconds (Garris & Wightman, 1994; Seamans & Yang, 2004; Wayment, Schenk, & Sorg, 2001). This means that although the dips are fast to reach the PFC, their destabilising effects would be slower and might only have an effect after a short delay. We suggest that the slow nature of the dopamine dip effects over the PFC might contribute to a delay between presentation of negatively conditioned stimuli and the following behavioural reaction.

To summarize, in the constructed model, information from negatively valenced emotional (conditioned) cues is propagated through affective areas (amygdala and the VTA) to higher cortical areas (PFC). This propagation induces a shift of the task set from the one currently imposed towards a different one that is more relevant to the conditioned stimulus. The induced competition between the current and the new tasks results in reduced activation of the enabled task representation (word reading or colour naming). This leads to slower processing of task-relevant stimuli and thus to slower responses, corresponding to the emotional Stroop effect. Importantly, because dopamine dips are slow to take effect in the PFC, task-set competition is only considered enabled between two consecutive trials rather than immediately. This leads to the between-trial slowdowns, as has been reported experimentally (Phaf & Kan, 2007). Because the conditioned stimulus-induced competition in the cognitive task set is the principal idea of the model, we term the model the *conditioned task-set competition* (CTC) account of the emotional Stroop effect. A description of the computational principles used in the study follows, with a detailed description of the model architecture and specification of model parameters.

Modelling methods and architecture

Modelling principles

We follow the connectionist principles of Cohen et al. (1990; and extensions, e.g. Botvinick et al., 2001; Wyble et al.,

2008). In brief, each unit in the model represents a population of neurons in the brain, characterised by a level of activation. The units are interconnected with each other, which roughly corresponds to (direct or indirect) white matter projections between neuron populations. Connection strengths from each unit mediate how much the unit's output influences activations of other units.

Each unit's activation is the running average of its inputs:

$$a_{i,t} = a_{i,t-1}(1-\tau) \; + \; i_{i,t}\tau \tag{1}$$

Here $a_{i,t}$ is activation of unit i at time t; $i_{i,t}$ is the total input to unit i at time t; and τ is the activation time constant.

Each unit's activation is modelled to be always close to or above zero (set to zero in case of a negative value). Each unit's complete input is the sum of weighted outputs of all units with incoming connections, together with the unit activation bias (external input in a trial):

$$i_{i,t} = \sum_j o_{j,t} w_{j,i} \; + \; b_i \tag{2}$$

Here $o_{j,t}$ - output from unit j at time t; b_i - bias / external input to unit i at current trial; $w_{j,i}$ - connection strength from unit j to unit i.

Finally, each unit's output is computed as its sigmoid-transformed activation:

$$o_{j,t} = \frac{1}{1 + e^{-\gamma_i\left(Sa_{i,t-1} \; - \; \theta\right)}} - d \tag{3}$$

In Eq. 3, $d = \frac{1}{1 + e^{\gamma_i \theta}}$ is the term used to force unit *output* to zero when unit *activation* is zero; and γ_i, θ, S - unit i output function parameters (γ_i is different between the processing, conditioned, and task-set layers).

During the course of a trial, unit activations are repeatedly updated until one of the units in the response layer achieves a prespecified threshold. The response is then considered to be achieved. This is a simplification from the original response mechanism of Cohen et al. (1990), which was based on evidence accumulation. Similarly simpler response mechanisms have been used successfully in other models (e.g. Wyble et al., 2008; Yeung et al. 2004). The number of update cycles of the unit activations is taken as representative of the trial response time. To relate the number of update cycles (RT_{cycles}) to a response time in milliseconds (RT_{ms}), the cycle numbers are translated in the following way:

$$RT_{ms} = RT_{cycles} * K \; + \; I \tag{4}$$

Here, K is the regression parameter (how many cycles correspond to a millisecond?) and I is the intercept parameter (how much time does it take for stimulus preprocessing and for response execution *outside* of the model?).

Response errors and variability in response times between trials have not been considered. Hence, performance of the constructed model is deterministic—no unit activation noise is included. This was also the case in the previous modelling account (Wyble et al., 2008).

Critically, the γ parameters above are representative of the gains of the unit output functions, such that higher γ parameter values result in sharper unit outputs. A higher γ value results in a sharp increase in the output when unit activation reaches a certain threshold. Lower values of the γ parameter result in a more linear relationship between the activation and output, with lower maximal output of the unit. An illustration of the effect of the γ parameter over the unit output function can be found in Appendix 1. A prominent neurobiological interpretation of the gain (γ) parameter for PFC has been proposed by Servan-Schreiber, Printz, and Cohen (1990), suggesting relevance to the levels of catecholamines (in particular, dopamine) effective over the activated units. This interpretation has been expanded upon and applied to explain dopaminergic deficiency aspects of schizophrenia in relation to cognitive deficits in the disorder (Cohen & Servan-Schreiber, 1993; Servan-Schreiber, Bruno, Carter, & Cohen, 1998; see also Braver, Barch, & Cohen, 1999). Recent neurophysiological evidence supports this interpretation of the dopamine effects over prefrontal neurons, specifically through D1 receptors (Thurley, Senn, & Lüscher, 2008). Drawing on these theoretical considerations, the γ_T (gain) parameter of the task set (PFC) units has been taken as representative of the levels of dopamine in the PFC in effect at D1 receptors in the current study.

Dopamine level in the PFC (gain of the task-set units, γ_T) is computed basing on the output levels of dopamine midbrain cells (VTA, represented by the reinforcement unit *PR* in the model):

$$\gamma_T = \gamma_{Tmin} \; + \; r_T o_R \tag{5}$$

Here, γ_{Tmin} represents the minimal output gain of the PFC units when no dopaminergic input is present, r_T is a dopamine-level scale parameter, and o_R is the output of the reinforcement unit (*PR*) at the end of the preceding trial. To simulate the slow time course of dopamine effects over the PFC, γ_T is only updated between every two consecutive trials and fixed during the course of a single trial.

Model architecture

The constructed model consists primarily of the two competing pathways for processing colours and words from preprocessed perceptual to response selection stages—as in the other models of the classical Stroop effect. The word-reading pathway (see Fig. 1, *IPw* and *PRw* connections) is stronger than the colour-naming pathway (see Fig. 1, *IPc* and *PRc*

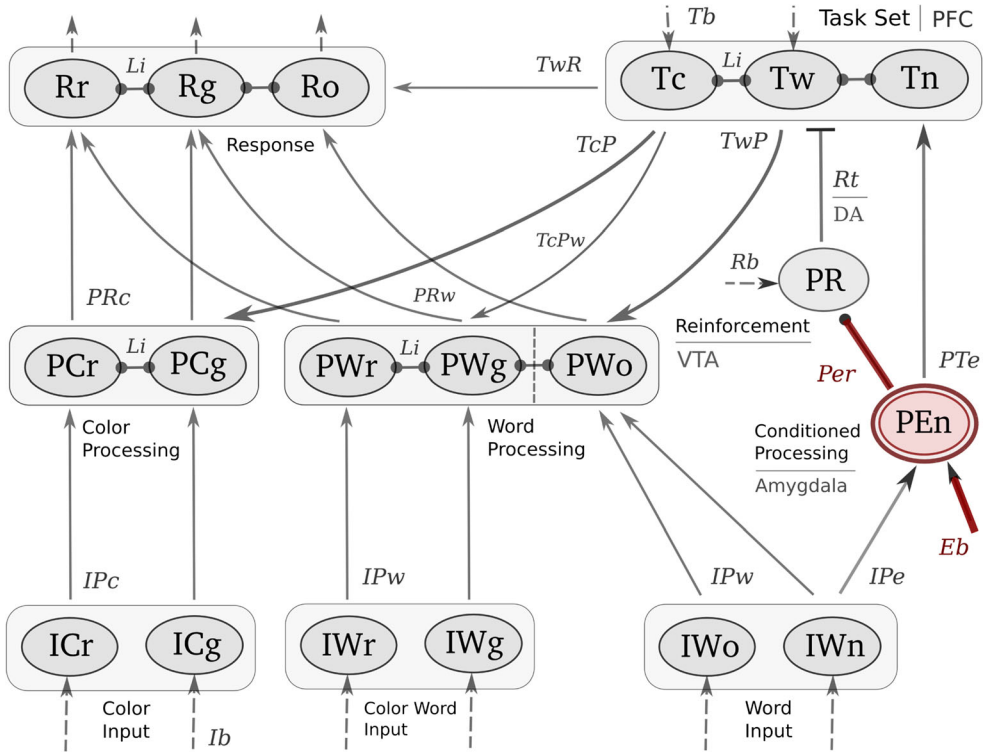

Fig. 1 Conditioned task-set competition model architecture. *Arrows* represent excitatory connections. *Circled arrows* represent inhibitory connections. *Solid lines* represent model connections. *Dashed lines* represent model inputs and outputs. Depressive mechanism (highlighted in *bold*): Increased tonic input to the conditioned processing unit (*amygdala*), and increased strength of inhibitory connection from the *conditioned processing* unit to the reinforcement prediction unit. *Abbreviations:* PFC = prefrontal cortex; VTA = ventral tegmental area; DA = dopamine

connections). The processing priority is imposed by the task-set activation (*Tc* or *Tw* unit), which supports activations in either the word reading, or in the colour naming pathway, to enable successful completion of the current task. Units within the classification, response and task-set layers have inhibitory connections between each other.

The classical Stroop effect is explained by a combination of mechanisms proposed in the model by Cohen et al. (1990) and its further extension (Herd et al., 2006). Colour-neutral and colour-incongruent words induce strong interference when the network identifies the colour, due to the stronger word-processing pathway connections. This results in slower responses at *both* neutral and incongruent conditions, compared to congruent. During *incongruent* trials, the colour-naming task unit further facilitates *all colour-word units* (through *TcPw* connection). This increases response competition compared to neutral trials and results in the highest slowdown at the incongruent condition—corresponding to the Stroop effect. This mechanism is similar to the account of Herd et al. (2006), where the colour-concept unit plays the role similar to the *TcPw* connection in the current model. Overall, the input, processing, response and task-set layers, and connections between them (see Fig. 1) are responsible for the classical Stroop effect, similar to the previous accounts (Cohen et al., 1990; Herd et al., 2006).

The third pathway is representative of processing conditioned stimuli. The pathway consists of two components: (1) Dopaminergic control of the task-set unit outputs (*IWn – PEn – PR – Tc / Tw / Tn* pathway) and (2) Conditioned information propagation to the task set in the PFC (*IWn – PEn – Tn* pathway). The *PEn* conditioned processing unit in the pathways is suggested to correspond to the amygdala. This unit in the model is only activated when processing negative words (*IWn* input unit active), but completely inactive otherwise for healthy controls. When activated, the unit exerts inhibitory control over the VTA-representative reinforcement unit *PR,* through connection *Per*. Inhibition of the *PR* unit is suggested to correspond to dopaminergic decrease signals (likely mediated by the intermediary RMTg structure). These dopamine signals propagate to the PFC (*Tc*, *Tw*, *Tn* units, connection *Rt*) and reduce output levels of those task-set units which are highly active—in order to facilitate competition (see Eq. 5 and Appendix 1). The negative task/concept representation in the PFC—*Tn* unit—then receives direct support from the *PEn* unit through connection *PTe*. This initiates competition between the task-set representations—*Tc* and *Tw* units against the *Tn* unit. The two mechanisms—dopamine level decrease and competition between tasks—contribute to decreased influence of the current task set over behaviour, and thus slower responses when negative emotional cues are present. Overall,

the conditioned processing (*IWn – PEn – Tn*) and the reinforcement (*IWn – PEn – PR – Tc / Tw / Tn*) pathways are responsible for the emotional Stroop effect.

It should be noted that during the course of a single negative-word trial, dopamine in the PFC (γ_T) stays at the same relatively high level due to the slow course of dopamine reuptake. Although the *Tn* negative concept unit receives conditioned information signals from the amygdala (*PEn* unit), it cannot become activated because of the high output of the active task set (*Tc* or *Tw* unit; see Appendix 1), and strong lateral inhibition. The same principle of high neural-gain mediated inhibition has been shown to focus information processing on highly representative or important features, and to limit learning (Eldar, Cohen, & Niv, 2013). Between two sequential trials, dopamine dip takes effect in the PFC, reducing the γ_T neural gain. Activations of units in the task set (*Tc*, *Tw* and *Tn*), as well as the conditioned pathway units (*PEn* and *PR)*, are carried over from the previous trial, mimicking a form of rudimentary working memory. Because of the lower neural gain, *Tn* unit is then able to become activated and to compete with the other task-set units. Both the reduced neural gain (γ_T) and the activated competing negative concept (*Tn* unit) decrease outputs of the highly active task-set units—*Tc* and *Tw.* This results in a delayed response in the trial *following* the negative-word presentation, accounting for the *slow* between-trial nature of the emotional Stroop effect.

Model constraints and specification

Performance constraints for the model were as follows. The model crucially had to produce correct responses at both colour-naming and word-reading tasks with colour (congruent and incongruent), neutral, and negative emotional words. For the classical Stroop task, the model was aimed to reproduce response times from the hallmark study of Dunbar and MacLeod (1984), Experiment 2, as in the previous connectionist accounts (Cohen et al., 1990; Herd et al., 2006; Wyble et al., 2008). For the emotional Stroop task, performance constraints were taken from McKenna and Sharma (2004), Experiment 3, and Algom et al. (2004), Experiment 2. In McKenna and Sharma (2004), a single negative word initiated a between-trial effect when presented in sequence with neutral words. Algom et al. (2004) have shown that the emotional slowdown predominantly occurs in blocks of trials, at both colour-naming and word-reading tasks. We selected these three datasets because they are highly representative of the classical and emotional Stroop reaction-time slowdowns. Accounting for these signatures enables a further investigation of deficits responsible for the effects of depression on reaction times. We describe below how the model parameters were specified and outline how the selected datasets were applied to constrain parameter values.

The constructed model has nine activation parameters—four input biases, one unit time constant, three output gain parameters, and one response threshold. These parameters were all fixed to either biologically or functionally reasonable values prior to the simulations, and are described in Appendix 1.

Apart from the activation parameters, the model contains 12 connection parameters (see Table 2 in Appendix 1). Three are responsible for the main processing pathway connections (*IPc, PRc, TS*), another four are responsible for task-set facilitation of the processing and response units (*TcP, TcPw, TwP, TwR*), one responsible for lateral inhibition between layer units (*Li*), and another four for task-set reinforcement and negative conditioned stimulus processing (*IPe, PTe, Per,* r_T). We describe specification of these parameters below.

First of all, the r_T connection represents dopaminergic fibers from the VTA to the PFC and was specified to enable a high neural gain of the task-set units when the reinforcement unit is highly active (with $r_T = 8$ and the fixed activation parameters, γ_T is close to the value of 8; see Eq. 5, Fig. 9 in Appendix 1). This represents relatively strong dopaminergic innervation of the PFC when performing a task. The lateral inhibitory connection strength (*Li*) was set to a relatively high value of 0.8—this means that once one unit is highly active within a layer, any competing unit must receive an input higher than 0.8 in order to produce any output. Together with a sufficiently high neural gain, this warrants strong lateral inhibition within layers. The *PRc* connection strength was tied to the response threshold (see Appendix 1) and specified to the value of 0.8. This connection was set to be sufficiently strong so that the maximal output of a processing unit could warrant its relevant response unit to cross the threshold and generate a response. The *IPc* and *TS* parameters concluded parametrization of the processing connections and were specified to sufficiently low values so that the model *could not* generate a response (cross the response threshold) without an active task-set unit (*IPc* = 0.5 and *TS* = 1.2; note that the *TS* parameter represents the training scale parameter and has to be above one; see Table 2 in Appendix 1).

We then explored and specified the four task-to-processing connections (*TcP, TcPw, TwP, TwR*) to replicate the *classical* Stroop effect. Specifically, the *TcP* and *TcPw* connections were set to replicate highest colour-naming reaction times at the incongruent condition, followed by neutral condition, and followed by congruent condition. *TwP* connection was specified to be strong enough to replicate approximately equal reaction times between the three conditions at the word-reading task. Strong *TwP* connection on its own could not account for the fast reaction times at the word-reading task. We hence added a connection from the word-reading task (*Tw*) unit to the response units (*Rr*, *Rg*, *Ro*, connection *TwR*). This is in line with the notion that the word-reading task is highly practiced and potentiates response activations when active. These four

task-set connection strengths were then manually tuned to best replicate 12 performance constraints: six reaction times (three for colour-naming and three for word-reading), and six related response correctness measures from Dunbar and MacLeod (1984, Experiment 2).

We finally specified the three remaining conditioned pathway connections (*IPe, PTe, Per*). *IPe* and *PTe* connection strengths were set to the value tied to the input bias of the task-set units (*IPe* = *PTe* = 1, since *Tb* is fixed to 1; see Appendix 1). This means that when a conditioned input is present (*IWn* unit active), its signal is propagated to the negative concept (*Tn*) unit in the task set, with the strength that matches input of the active task unit. This is considered to represent strong conditioned-stimulus neural signals from the amygdala to the PFC, which should enable lateral competition. Without the *Per* connection, however, conditioned signals are propagated to the PFC but cannot activate the *Tn* negative concept unit due to lateral inhibition. The *Per* connection strength was finally specified to replicate the *emotional* Stroop effects. Specifically, it was selected to best replicate *three* slowdown effects with negative words: the between-trial slowdown, when a single negative word is presented in a sequence with neutral words (McKenna & Sharma 2004, Experiment 3), and the two slowdown effects when negative words are presented in trial blocks at colour-naming and at word-reading tasks (Algom et al., 2004, Experiment 2). Additional constraint came from the fact that the model had to produce correct responses with negative words, despite the task-set destabilisation.

Overall, 17 behavioural performance data points were applied—six reaction times and six response correctness measures at the classical Stroop task (Dunbar & MacLeod, 1984, Experiment 2), three negative-word slowdown effects (McKenna & Sharma 2004, Experiment 3; Algom et al., 2004, Experiment 2), and another two response correctness measures—when negative words are presented in blocks at colour-naming and at word-reading tasks. (The resulting set of connection parameter values can be found in Table 3 in Appendix 1.)

Depression modelling

The second main aim of this study was to investigate the mechanisms of the increased response times and the emotional Stroop effect in depression. To this end, following the principles of computational psychiatric modelling (Maia & Frank, 2011), we have investigated alterations in the constrained model. Two main criteria have been applied to identify plausible depression mechanisms. First, the alterations had to reproduce the reported behavioural deficits—increased response times in negative, incongruent, and neutral trials (Epp et al., 2012). Second, the alterations were constrained to be relevant to the most prominent reported neural abnormalities in depression. A more detailed discussion of the second constraint is as follows.

First, significant neuroimaging evidence indicates alterations in the amygdala as one of the key features of depressive disorder (e.g. Drevets, 2003; Whalen et al., 2002). A recent meta-analysis of neuroimaging studies strongly supported amygdala hyperactivation in response to negative affective stimuli (Hamilton et al., 2012). Higher depressive amygdala activity has also been observed specifically in response to *conditioned* cues predicting occurrence of aversive pictures (Abler, Erk, Herwig, & Walter, 2007). These results indicate stronger neural processing of negatively valenced conditioned information in the amygdala in depressive disorder.

The amygdala deficit in depression is in line with the prominent theoretical reviews, as we have reviewed in the Background section. To summarize, Mayberg (1997), DeRubeis et al. (2008), Disner et al. (2011), and Roiser and Sahakian (2013), in complementing reviews, have suggested that limbic brain areas, including the amygdala, are hyperactive in depression. Higher cortical areas, including the DLPFC, are, on the other hand, hypoactive.

With regard to dopaminergic neurotransmission, as we have overviewed previously, several reviews have suggested deficits mainly in the mesolimbic dopamine pathway (Dunlop & Nemeroff, 2007; Nestler & Carlezon, 2006; Pizzagalli, 2014). We consider dopamine transmission a contributory factor for generating the emotional Stroop effect in the current model, and hence suggest that dopamine deficits might be relevant for the effects of depression at the task.

Drawing on the evidence overviewed above, we hypothesized that a combination of parameter changes in the conditioned processing (amygdala), reinforcement prediction (dopamine release) and task-set (PFC) units and connections could account for the pattern of depressive performance at the Stroop tasks. Appendix 2 outlines the set of parameters investigated to reproduce the depressively abnormal Stroop performance. Our aim has been to identify the *simplest* combination of parameter changes which closely replicates the pattern of behavioural deficits, and is most consistent with the neural mechanisms of depressive illness. We have hence given priority to the combinations which involved the lowest number of parameters and included at least one of the parameters governing the *PEn* conditioned processing unit activation. This constraint was aimed to mimic the well-supported hyperactivity of the amygdala.

Modelling results

Classical Stroop effect account

Similar to the previous connectionist accounts, the parametrized model accounts relatively well for the classical Stroop

effect (Cohen et al., 1990; Herd et al., 2006). Figure 2 shows performance of the model compared to the experimental Stroop effect. A possible limitation is in the quantitative match of colour-naming response time in the congruent condition—the model predicts a shorter response time than reported experimentally. This limitation, however, is also characteristic of the previous models (Cohen et al., 1990; Herd et al., 2006; Wyble et al., 2008). To additionally confirm the functional significance of the task-set units, the model was run with these units disabled (input bias set to zero)—this resulted in incorrect model performance with no response generated at neutral and incongruent trials at either of the tasks.

To check if the model is able to account for the classical Stroop neuroimaging results of Banich et al., (2000, 2001), average colour and colour-word processing unit activations across all trial cycles in each condition during the colour-naming task were computed. The results are illustrated in Fig. 3 and show that the pattern of neuroactivations predicted by the model qualitatively matches the neuroimaging reports. Higher activation can be observed in the word-processing layer (verbal cortical area) in incongruent, compared to congruent and neutral trials. This is in line with the model of Herd et al. (2006, Fig. 3). The model can be considered a simplification of the TEB account by Herd and colleagues. In particular, the general concept of colour from the TEB account is replaced by the connection from the colour-naming task-set unit to the two colour-word units in the word-processing layer (Fig. 1, connection *TcPw*). This is consistent with the notion that the general concept of colour, as suggested by Herd et al., is recruited as part of the colour-naming task set in the constructed model.

To additionally confirm utility of the *TcPw* connection, the model was simulated with this connection disabled. The model still produced correct responses, but no longer reproduced the classical Stroop effect—response times at incongruent and neutral conditions appeared similar. The neuroactivations effect illustrated in Fig. 3 could no longer be reproduced—average activations of the word-processing units appeared similar in the neutral and incongruent conditions. These results

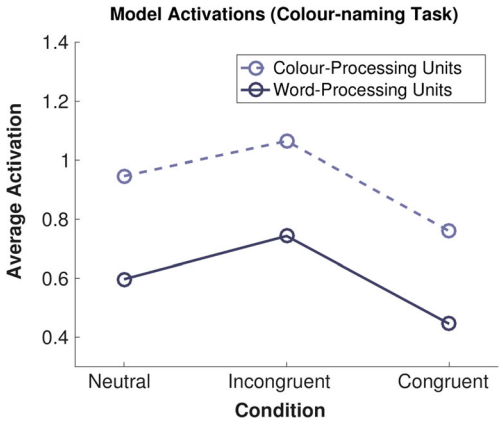

Fig. 3 Modelled classical Stroop colour-naming task neuroactivations in word- and colour-processing units

support the theory of Herd et al. (2006), which suggests that task-related information processing is facilitated in all dimensions, including those which may not be relevant for the task (i.e. colour-related information in verbal areas).

Emotional Stroop effect account

The specified model can account for the experimentally reported slow emotional Stroop effect, occurring at both colour-naming and word-reading tasks.

To replicate the carry-over slow emotional Stroop effect, we simulated the model in accordance with the experimental conditions of McKenna and Sharma (2004). An array of consecutive colour-naming trials was executed, where the first trial word was negative, while the following trial words were neutral. Results of the sequence simulation may be seen in Fig. 4. Experimental response times were extracted and replotted from McKenna and Sharma (2004, Experiment 3, Fig. 1). As can be noted from Fig. 4, the model accounts relatively well for the between-trial slowdown effect of a single negative word presentation. The modelled colour-naming response time at the trial immediately following the negative-

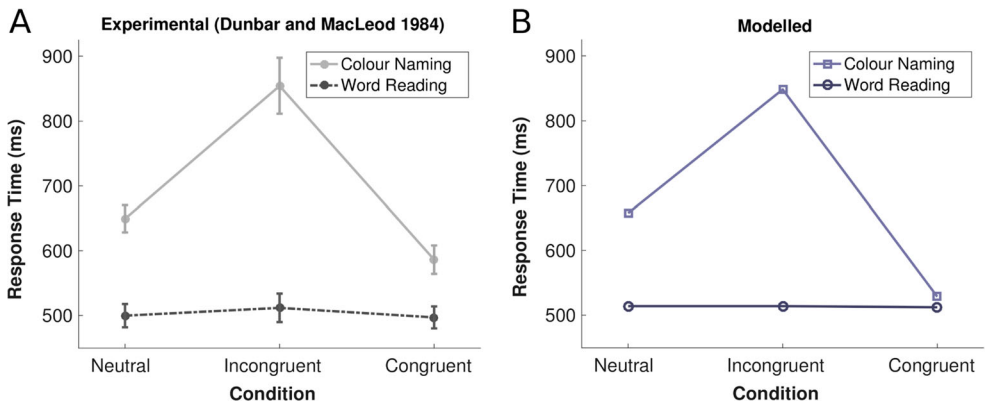

Fig. 2 Classical Stroop interference effect, as reported in Dunbar and MacLeod (1984, Experiment 2, Fig. 3) (**a**), and modelled replication (**b**). Experimental response times and standard errors extracted and replotted. Model regression and intercept parameters: $K = 1.82$ ms/cycle, $I = 398$ ms

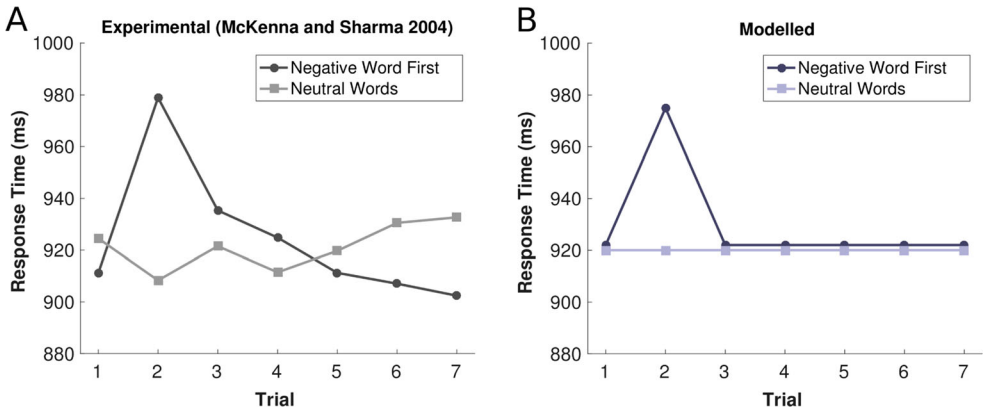

Fig. 4 Slow emotional Stroop (sequence) effect, as reported in McKenna and Sharma (2004, Experiment 3, Fig. 1) (**a**), and modelled replication (**b**). Experimental response times extracted and replotted. Only second trial difference between negative-word sequence and neutral-word sequence response times is significant in both the experimental data and the modelled replication. Regression and intercept parameters: $K = 3.06$ ms/cycle, $I = 483$ ms

word trial (first in the sequence) is significantly slower than at the trials from the sequence with no negative words (Fig. 4b, Trial 2, 975 ms vs. 920 ms).

Algom et al. (2004) showed that when negative and neutral word trials are presented in blocks, the emotion-related response delay effect extends not only to the colour-naming task but also to the word-reading task. Depending on the experimental conditions, the delay observed for blocked negative words at word reading appeared just as high as at colour naming. To replicate these results, the model was simulated at colour naming and word reading with the trials presented in blocks. The resulting mean simulated response times compared to the original experimentally reported data are presented in Fig. 5. The specified model accounts well for the effect of negative words at colour naming in blocks of trials. For word reading, the model accounts for the significant slowdown with negative words, but the predicted effect is slightly lower than experimentally reported (Fig. 5, word-reading task, approx. 25 ms modelled vs. approx. 35 ms experimental). The model thus replicates the blocked emotional Stroop effect, but predicts a higher magnitude of the effect at the colour-naming task compared to word reading. This is in a slight contrast to the results reported by Algom et al. (2004), which indicate comparable effects in both tasks.

To summarize, the model accounts well for the emotional Stroop effect both in sequence of mixed words (Fig. 4; McKenna & Sharma, 2004), and when negative words are presented in blocks (Fig. 5; Algom et al., 2004). The model thus captures both the emotional reaction-time slowdown and its predominantly slow between-trial nature (Phaf & Kan, 2007).

Depression modelling results

Alteration in no single model parameter, from those selected as relevant for depression (Appendix 2), could account for the entire pattern of depressive deficits. During further exploration, we assumed that an alteration in at least one of the three

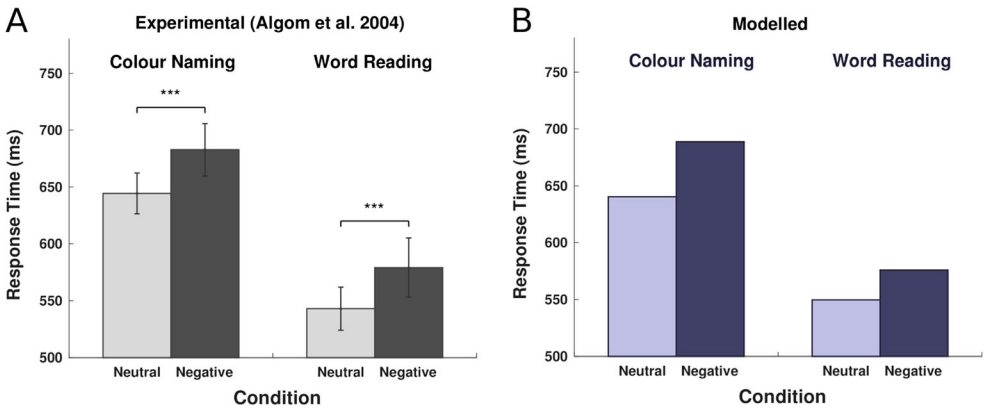

Fig. 5 Blocked emotional Stroop effect, as reported in Algom et al. (2004, Experiment 2, Fig. 2) (**a**), and modelled replication (**b**). Experimental response times and standard errors were extracted and replotted. *Triple stars* indicate highly significant difference ($p < .001$). Colour-naming was reported significantly slower compared to word-reading (not shown in figure). All differences between model performance statistics are significant (no variability was modelled). Regression and intercept parameters: $K = 1.15$ ms/cycle, $I = 476$ ms

parameters governing activation of the conditioned processing unit (*IPe,* conditioned unit input connection strength; *Eb*, conditioned unit input bias; or the conditioned unit output gain) must be present in depression. Increase in either of these parameters could be considered representative of hyperactivity of the amygdala.

Touples of parameter alterations were explored with the above constraint. Results revealed one simple plausible combination involving alteration of two parameters: increased tonic activity in the conditioned processing unit (*Eb* increase from zero to a moderate value), and increase in inhibitory connection strength between the conditioned processing unit and the reinforcement unit (*Per* connection strength increase). This corresponds to moderate baseline hyperactivity of the amygdala and stronger inhibition of mesocortical dopamine release in depression. Specific details of the depressive parameter alterations can be found in Appendix 2. An illustration of the identified depression mechanism is highlighted in bold in Fig. 1.

The identified depressive mechanism (*Eb* input increase; *Per* connection strength increase) could generally replicate the slowdowns at the Stroop tasks. Highest slowdown is observed at the negative condition, followed by the incongruent condition, and then neutral condition (see Fig. 6). This is in line with the meta-analytic results by Epp et al. (2012). Epp and colleagues reported a highly significant Hedges' *g* value of 0.98 for the effect of depression in the negative-word Stroop condition (basing on 19 studies), followed by 0.86 for incongruent condition (basing on 14 studies), and 0.81 for neutral condition (basing on 17 studies). Hedges' *g* is a standardized effect size measure computed by normalizing the difference between sample means with a corrected measure of the pooled standard deviation (Durlak, 2009). For qualitative comparison of effect sizes between conditions, the simple absolute mean difference between participant samples (depression and control) could be considered equivalent to the Hedges' *g* (or other standardized effect sizes)—drawing on an assumption that the pooled standard deviations are very close or similar between all conditions. This could be considered a generally reasonable assumption for experimental conditions of the Stroop and emotional Stroop tasks. In terms of condition mean differences between samples, the modelled depression mechanism generated the following simulated effects (in cycles): 108 in negative-word condition; 84 in incongruent condition; followed by 51 in neutral condition, and 21 in congruent condition. These results are *qualitatively* in line with the meta-analytic report by Epp et al. (2012), but also predict a small effect at the congruent condition.

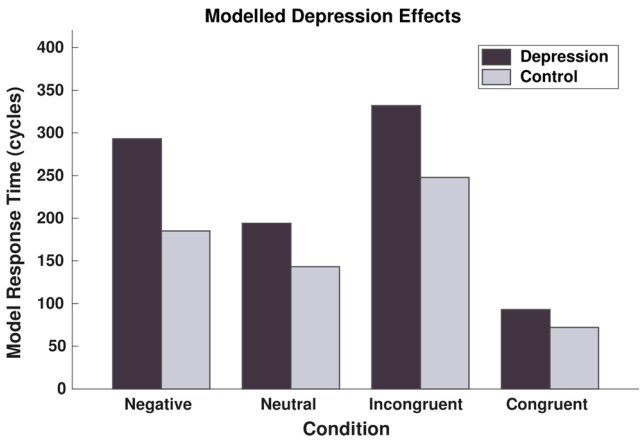

Fig. 6 Depression mechanism simulation results. Introduction of the depressive mechanism leads to—in the order of absolute effect size—slower responses in the negative condition, followed by incongruent condition, followed by neutral and, finally, congruent conditions

Figure 7 illustrates how the model performance compares against the experimental data for the classical Stroop task in depression. Holmes and Pizzagalli (2008) have reported that depression was associated with significantly slower response times at the incongruent, compared to the congruent condition. The depression model can replicate these results. The quantitative aspect of this fit, however, should not be taken too strongly. The authors only report behavioural results for two experimental Stroop conditions in their study: congruent and incongruent. With only two metrics and two parameters inferred for translating simulated response times from model cycles to milliseconds (regression coefficient and intercept; Eq. 4), there is a risk of overfitting the regression model. These results should be taken as an illustration of a qualitative fit of the model to the depressive performance at the classical Stroop task as well as an illustration of the *potential* of the model to quantitatively fit depression behavioural data.

With regard to the emotional Stroop task, the model performance was compared to the behavioural results reported by Mitterschiffthaler et al. (2008). Mitterschiffthaler and colleagues have not reported a significant difference between depression and control performance at the neutral-word condition, despite a trend towards slower responses. Depressed participants in the study have shown significantly slower responses to negative words, compared to controls. Figure 8 illustrates performance of the model compared to these results. The model qualitatively replicates the response time increase in the negative-word condition in depression. As previously, these results should not be taken as a strong claim of a good quantitative fit to the depressive behavioural data, due to only two reported control experimental conditions (negative word and neutral word) modelled to derive the regression parameters. Compared to the results by Mitterschiffthaler and colleagues, the model predicts a significant depressive slowdown at the neutral condition—in line with the meta-analysis results (Epp et al., 2012).

To check how each of the two depressive alterations contributes to the behavioural effects, we simulated them separately. Results revealed that both deficits (hyperactive

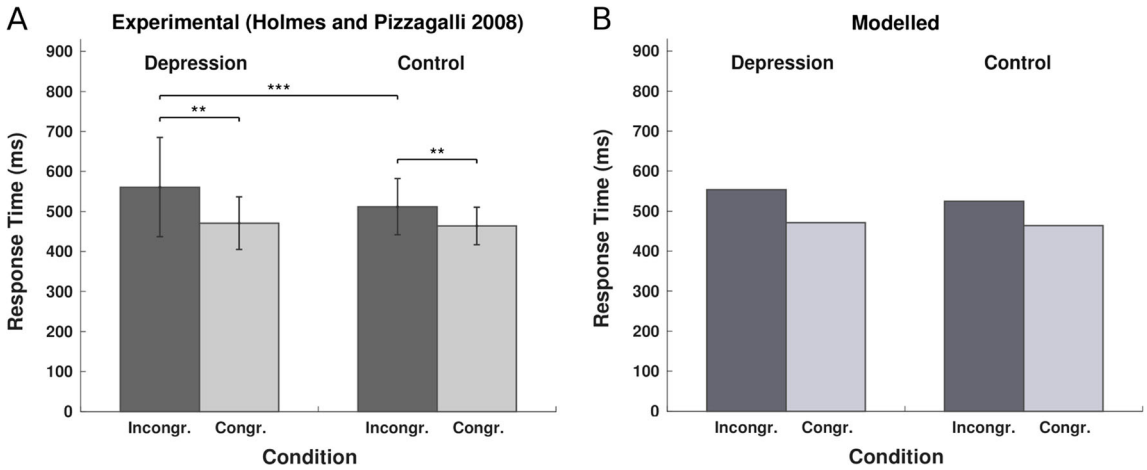

Fig. 7 Classical Stroop effect in depression, as reported in Holmes and Pizzagalli (2008, Table 1a) (**a**), and modelled replication (**b**). *Double stars* indicate high significance ($p < .01$), *triple stars* indicate highest significance ($p < .001$). All modelled condition response time differences are significant (no variability has been modelled). Regression and intercept parameters: $K = 0.35$ ms/cycle, $I = 439$ ms

amygdala and increased inhibitory influence of the amygdala over the VTA) are responsible for the increase in response times at the neutral and incongruent conditions. Output generated by the tonically hyperactive *PEn* unit (amygdala) propagates to the task set and results in weaker influence of the *Tc* colour-naming task unit over task-related processing (*PCr* and *PCg*), resulting in response delays. This is, however, only possible when the *Per* connection strength is increased (DA release from the VTA is sufficiently inhibited), which warrants decreased task-set output gains and enables the amygdalar signal to propagate. Only the increased *Per* connection strength, on the other hand, appeared responsible for the higher response times at the negative-word colour-naming trials, with little effect from the tonically hyperactive *PEn* unit. The stronger *Per* connection decreases DA release specifically when negative words are present, and thus increases influence of negative words over the task-set stability.

Exploration of the parameter space revealed that several other simple deficit combinations could also replicate the depressive behavioural effects. Specifically, hyperactive amygdala and either the decreased reinforcement unit output gain, the decreased reinforcement unit input bias, or the decreased dopaminergic connection from reinforcement to task set, could also replicate the effects. These alternative mechanisms, however, all appeared less biologically relevant to depression than the main identified combination. Details of the alternative mechanisms are described in Appendix 2. We briefly overview them and highlight evidence favouring the main identified combination in the Discussion section.

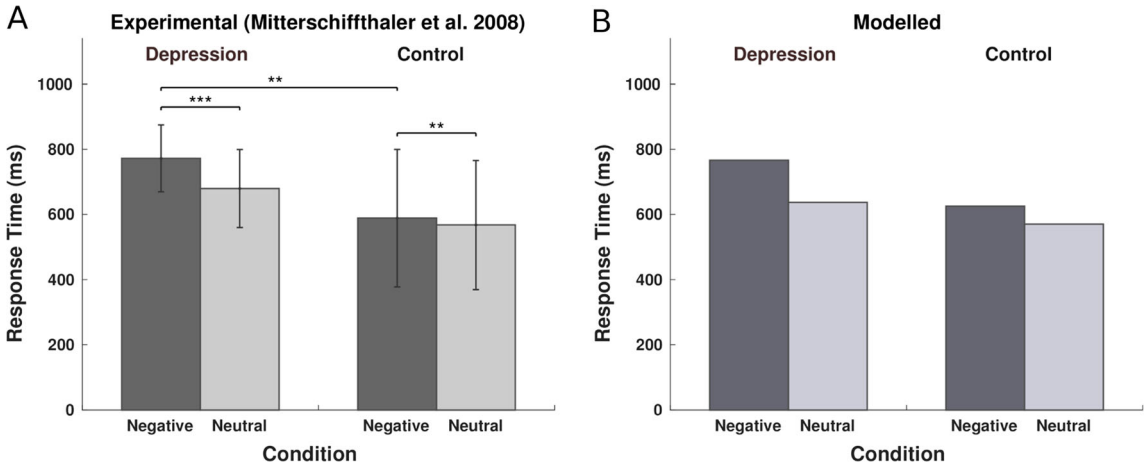

Fig. 8 Blocked emotional Stroop effect in depression, as reported in Mitterschiffthaler et al. (2008) (**a**), and modelled replication (**b**). Experimental response times and standard errors extracted and replotted. *Double stars* indicate high significance ($p < .01$), *triple stars* indicate highest significance ($p < .001$). All modelled condition response time differences are significant (no variability has been modelled). Regression and intercept parameters: $K = 1.31$ ms/cycle, $I = 383$ ms

Discussion of results

Modelled mechanisms of the emotional Stroop effect

Neurobiological correlates

The constructed model is largely neurobiologically driven and is consistent with the existing neuroimaging evidence indicating activation of the amygdala at the emotional Stroop task (Isenberg et al., 1999), as well as evidence of the amygdala activation in response to negative emotional words (e.g. Hamann & Mao, 2002; Naccache et al., 2005; Straube et al., 2011).

Whalen et al. (1998) and Mohanty et al. (2007) reported activation of the rACC at the emotional Stroop trials. Our model does not explicitly account for the rACC activation; however, two interpretations can be given. Namely, the rACC could either be involved directly in generation of the emotional Stroop effect, or recruited as a compensatory mechanism to maintain correct performance in the face of affective distraction (as suggested e.g. in Mohanty et al., 2007). In the first case, the rACC could in part be represented by the *Tn* negative-concept unit in our model—providing competition with the task representation (*Tc* unit) in the DLPFC, and generating the slow effect. The second case—suggestion of a compensatory role of the rACC—is consistent with the evidence of involvement of the rACC in resolving emotional conflict and inhibiting the amygdala (e.g. Egner et al., 2008; Etkin et al., 2006). In the second case, the rACC would not contribute directly to generating the emotional Stroop effect and could hence be left safely outside of the scope of our model.

We suggest that propagation of conditioned information to higher cortical areas results in competition between representations of the current task set and the concepts related to conditioned information. We do not specify where exactly in the PFC this competition might take place; however, some existing theories provide an indication. Badre (2008) has proposed that the PFC is organized in a rostro-caudal hierarchy—with more anterior regions containing progressively more abstract representations of contexts and goals. In our model, conditioned information competition could occur in the more rostral, abstract concept-related areas of the PFC, with subsequent destabilizing effects over the more caudal prefrontal areas, including the DLPFC, which hold specific behavioural task-set representations. OFC has been suggested to extract and store valuation associated with conditioned information (e.g. Holland & Gallagher, 2004; Frank & Claus, 2006). MPFC has been suggested to have a role in selection of appropriate actions (e.g. Frank, Cohen, & Sanfey, 2009). Drawing on the reported connectivity of the amygdala (Carmichael & Price, 1995; Ray & Zald, 2012), it is possible that conditioned information is primarily propagated to the OFC and the MPFC, where it triggers conflict between higher-level context representations. Effects of this conflict may then propagate to more caudal task-related areas, which results in task deactivation and behavioural slowdowns.

Computational modelling accounts

The earliest connectionist account of the emotional Stroop effect has been proposed by Matthews and Harley (1996). The authors suggested that the slowdown effect arises because of excitatory facilitation of affective information processing, which results in the task-related response interference—similar to the classical Stroop effect. The early model by Matthews and Harley does not account for the predominantly slow intertrial nature of the emotional Stroop effect (McKenna & Sharma, 2004; Phaf & Kan, 2007). No interpretation is considered by the authors as to how the excitatory facilitation of the emotional and threat-related information might be implemented in the brain. This is in contrast to our model, which explains the slow effect and is neurobiologically driven.

From the perspective of motivational significance of the *slow* emotional Stroop effect, our model is conceptually consistent with the previous account by Wyble et al. (2008). Wyble et al. suggest that the slow effect occurs because of deallocation of cognitive resources away from the current colour-naming task in order to deal with the negative-word signalled threat. Our model supports this notion; however we suggest that cognitive task-set resources are more specifically *reallocated* towards processing negative material, rather than simply *freed* from all tasks (decreased cognitive control) as suggested by Wyble and colleagues.

The distinct advantage of our model over the previous accounts is that we specifically consider the neural mechanisms of conditioned stimuli processing in generating the emotional Stroop effect. Wyble and colleagues suggest a neurobiological interpretation which implicates competition between the rACC (emotional monitoring) and the dACC (cognitive control) in generation of the *slow* effect. The authors note that this is disputable since little direct evidence of such competition has been reported. In contrast, we suggest that the *slow* effect is generated due to propagation of negative conditioned information from limbic (amygdala) to higher cognitive areas (PFC), which is generally neurobiologically plausible.

Our computational model is constructed with several simplifications which are worth mentioning. First of all, we do not precisely model phasic (time-limited) dopaminergic dip signals observed in neurobiology. The model rather presents a simplified notion of amygdala-induced decrease in dopaminergic neurotransmission. Existing investigations show that phasic dopaminergic neurotransmission is contingent upon presentation of the CS with the fast-onset dips lasting over a second (e.g. Mileykovskiy & Morales, 2011; Oleson, Gentry, Chioma, & Cheer, 2012). Response times in the emotional

Stroop task are usually below one second. For simplicity, we have modelled phasic dopamine dips to have an effect over an entire following trial, and left a more detailed account of these signals safely outside of the scope of our current model.

In the study, the single set of model parameters was able to account for both the classical and emotional Stroop effects. It can be noted, however, that regression and intercept parameters of our model (Eq. 4) vary between the simulated experiments (Figs. 2, 4–5, 7–8). This accounts for cognitive processing differences in different experimental conditions. Differences in the regression parameter K are representative of differences in the speed of processing within the connectionist architecture (Fig. 1), while differences in the intercept parameters I are representative of the different amounts of time necessary for visual preprocessing and motor mechanics. Variability in the intercept parameters between conditions is generally plausible, with values ranging between 380 and 490 ms across the five modelled experiments. Regression and intercept variability is also characteristic of the previous account of Wyble et al. (2008). Although we specified each parameter with sensible constraints, reasonable variations are highly plausible and would likely correspond to individual biological or behavioural differences.

Our model serves mainly as a proof of principle that neural mechanisms of conditioned stimuli processing can account for the behavioural emotional Stroop effect. We suggest that emotional words serve as negatively conditioned stimuli, and hence that mechanisms of conditioned stimuli processing could be responsible for the reaction-time slowdowns. Because our model is a simplified computational bridge between the neural mechanisms and behaviour, the model fits provide proof that, in principle, this is possible. We believe that these results offer a new perspective on the mechanisms behind the emotional Stroop effect, which could guide future investigations with both healthy and clinical participants.

Experimental predictions

Our theoretical account makes several predictions. We consider negative emotional words a case of conditioned stimuli. This implies that response delays at the emotional Stroop task could be reproducible when negative words are replaced with experimentally aversively conditioned stimuli. To test this prediction, experimental participants could first undergo a conditioning procedure—with neutral stimuli paired with aversive shocks (e.g. as in Raio, Carmel, Carrasco, & Phelps, 2012), or paired with instrumental responses to avoid shocks. These same (now conditioned) stimuli could then be used instead of words at the colour-naming task. We predict that response delay effects should occur with both aversively conditioned stimuli and with negative words.

Second, we suggest that the emotional Stroop delay effect depends crucially on the dopaminergic decrease signals reaching the PFC. We thus predict that tonically increasing dopamine levels in the PFC—for example, through administration of dopamine degradation inhibitor tolcapone (e.g. as in Kayser, Allen, Navarro-Cebrian, Mitchell, & Fields, 2012)—should decrease the impact of the dip signals and counteract the effect – either decreasing or eliminating occurrence of the reaction time slowdowns.

Finally, we suggest that the slow emotional Stroop effect is dependent on propagation of conditioned information to the prefrontal cortical areas—likely the MPFC and the OFC—to facilitate reallocation of cognitive resources from the current task. We thus predict that negative-word time-locked excitatory stimulation of these areas, through application of anodal transcranial direct current stimulation (TDCS; e.g. Bellaiche et al., 2013), should enhance representations of conditioned information and increase the delay effects at the emotional Stroop trials. Inhibitory cathodal stimulation, should, on the other hand, impair propagation of conditioned information and ameliorate the delay effects.

Modelled mechanisms of depressive task-set interference

In the current investigation of mechanisms at play at the Stroop tasks in depression, we broadly followed the *deductive* approach, as termed by Maia and Frank (2011) in their review of computational psychiatry and neurology modelling methods. We constructed a connectionist model of normal performance at the Stroop tasks and specified it to explain the hallmark behavioural findings (Figs. 2–5). We then investigated alterations which are most relevant for depressive disorder and introduced a simple mechanism which could generally account for the effects of depression at the tasks (Figs. 6–8)—hyperactive amygdala and stronger functional inhibitory influence of the amygdala over the VTA dopamine neurons. To our knowledge, this is the first explicit mechanistic theoretical account of depressive performance at the classical and emotional Stroop tasks. Given the reported correlation between the symptom severity and the response time effects in depression (Epp et al., 2012), these mechanistic deficits could be highly relevant to the symptoms of the disorder.

Several neuroimaging studies have linked depressive hyperactivity of the amygdala to rumination (Cooney, Joormann, Eugène, Dennis, & Gotlib, 2010; Mandell, Siegle, Shutt, Feldmiller, & Thase, 2014). In our model of depression, tonic amygdalar hyperactivity triggers persistent competition for resources in the PFC between conditioned negative information and task representations. This could be interpreted as representative of ruminative processes in the disorder. The novelty of our investigation is therefore that we suggest a mechanistic link between depressive ruminative processes and executive deficits at the classical Stroop task. Several previous behavioural studies also support the assertion that depressive rumination might be linked to executive deficits

(e.g. Jones, Siegle, Muelly, Haggerty, & Ghinassi, 2010; Levens, Muhtadie, & Gotlib, 2009; Watkins & Brown, 2002).

When we explored the depression-relevant parameter space we were able to replicate the depressive reaction-time effects with other simple combinations that included the amygdalar hyperactivity (see Appendix 2). Existing experimental evidence, however, favours the main selected mechanism. Increased gain of the reinforcement unit (first alternative mechanism; see Appendix 2) could be considered representative of higher responsiveness to rewards in the dopaminergic system. Existing evidence, however, indicates that reward encoding is decreased in main dopamine targets including the OFC and the striatum, while the dopamine transmission is likely decreased (Pizzagalli, 2014). Significantly decreased dopaminergic connectivity from the VTA to the PFC (second alternative mechanism, Appendix 2) would indicate extensive white-matter tract abnormalities between the two regions; however, only limited evidence of such deficits in depression has been reported (see Bracht, Linden, & Keedwell, 2015 for review). Finally, decreased baseline VTA activity during task performance (third alternative) could be a plausible alternative mechanism mediating dopamine deficits in depression. Available experimental evidence, however, indicates that decreased dopamine transmission is likely mediated by an active inhibition process rather than internal VTA factors. Tye et al. (2013), for example, directly optogenetically inhibited midbrain dopamine neurons, which reproduced depression-related behaviours in rats. Tanaka et al. (2012) reported that attenuation of the VTA dopamine neurons in depression-susceptible mice is likely mediated by enhanced VTA inhibitory inputs, due to increased levels of a specific bioactive lipid—prostaglandin E2. Chang and Grace (2014) have also reported decreased activity of dopamine neurons in a rat model of depression. Crucially, this deficit was reversed by pharmacologically attenuating activity of either the ventral pallidum (VP) or the basolateral amygdala (BLA). Further, pharmacological activation of the BLA decreased dopamine neuron activity in control rats. Chang and Grace suggested that depressive behaviour could be mediated by inhibition of the VTA dopamine neurons by the BLA, mediated by the intermediary VP structure. Altogether, these experimental results provide a compelling argument favouring the main selected depression mechanism over the three possible alternatives.

Mitterschiffthaler et al. (2008) have reported stronger activation of the rACC in depression during performance of the emotional Stroop task. The authors suggested that hyperactive rACC could represent a compensatory mechanism—suppressing emotional processing in the amygdala. If this is indeed the case, we do not explain hyperactivation of the rACC at the task since we only consider mechanisms which are directly involved in generating the task interference effects. Alternatively, the rACC might be involved in propagation of negative conditioned information towards higher cognitive processing—in this case, stronger activation of the negative concept (*Tn*) unit due to tonic conditioned signals in our model might in part be representative of the rACC hyperactivity.

In the model we focus on the effects of negative words because negative and threat words are most widely used in the emotional Stroop paradigm (Phaf & Kan, 2007). Epp et al. (2012), however, also reported a significant reaction time slowdown with positive words in depression (Hedges' *g* of 0.87). This effect was higher than with neutral words (Hedges' *g* of 0.81), but lower than with negative words (Hedges' *g* of 0.98). Although we do not provide an explicit account, we suggest that this effect could be due to the same neural mechanisms. Specifically, positive words could still be processed by the amygdala (as conditioned stimuli), but would not trigger dopamine dips, which would limit their behavioural effect in healthy participants. In depression, decreased dopamine transmission would enable the positive-word signals to propagate to the PFC, which would result in task-set competition and response delays. Because of a lack of added dopamine dip signals with positive words, however, these delays would be lower than with the negative words, but higher than with neutral words, as reported in Epp et al. (2012).

Drawing on the modelled depression mechanism, we make several experimentally testable predictions. We suggest that the amygdala exhibits stronger inhibitory functional influence over the VTA in depression. This is directly testable through dynamic causal modelling (DCM)—a technique used successfully to investigate functional interactions between brain regions (Etkin et al., 2006; Friston 2009; Friston, Harrison, & Penny, 2003). Depressed patients should exhibit stronger inhibition of the VTA by the amygdala compared to controls—either at rest, or during task performance in the scanner. Because of the tonically inhibited dopamine transmission, the model also predicts that the depressed participants should exhibit a small *fast* (same-trial) emotional Stroop effect alongside the increased *slow* between-trial response delay, due to potentiated processing of negative material.

Conclusion and further investigations

We have proposed a novel integrative model of the mechanisms at play when generating the classical and emotional Stroop effects. Our theory is based on the novel interpretation of emotional words as a specific case of conditioned stimuli. We grounded the model with aspects of neurobiology involved in conditioned stimuli processing. We suggest that the slow between-trial emotional Stroop effect is mediated by dopamine decrease signals, which reach the PFC and enable the amygdala-initiated competition for resources. We suggest that depressive deficits in the Stroop tasks might be caused by the hyperactive amygdala and the increased functional inhibitory influence of the amygdala over dopaminergic neurotransmission. Because of the

reported correlation between depression severity and task performance (Epp et al., 2012), we suggest that these proposed mechanisms might be highly relevant for understanding depressive illness. We believe that these results offer a new perspective on the mechanisms of the emotional Stroop effect in health and in depression, which could lead future investigations. We offered several experimental predictions, testable though behavioural, cortical stimulation, pharmacological and neuroimaging methods—future studies will test and prove or disprove aspects of our theory and the suggested depression mechanisms.

One particular avenue for both theoretical and experimental future investigations is the role of the rACC at the emotional Stroop task. Existing neuroimaging studies have shown activation of the region when performing the task (Bush et al., 2000; Mohanty et al., 2007; Whalen et al., 1998), and Mitterschiffthaler et al. (2008) have reported hyperactivity at negative-word trials in depression. Future studies should better explain the functional role of this region and might indicate whether its hyperactivation at negative word trials is relevant for symptomatology of depression.

Acknowledgements The current work was funded by Grants EP/F500385/1 and BB/F529254/1 for the University of Edinburgh, School of Informatics, Doctoral Training Centre in Neuroinformatics and Computational Neuroscience from the UK Engineering and Physical Sciences Research Council (EPSRC), the UK Biotechnology and Biological Sciences Research Council (BBSRC), and the UK Medical Research Council (MRC).

Appendix 1 | Model parametrization

Model unit activation parameters are briefly described in Table 1. Input bias *Ib* and *Tb* parameters are constrained to the value one at the trial-relevant task unit and the trial-relevant input units. The canonical value of one follows from the fact that output of a single unit is constrained between zero and one (see Fig. 9). *Eb* is always zero for control (healthy) participant simulations. All other units do not receive external inputs from outside of the model. Time constant of unit activations has been fixed to the value $\tau = 0.025$ for all units, although this is not principal for the model performance. Response threshold has been constrained to $R_{Th} = 0.75$, representing a high response unit activation necessary for response generation. All input, processing and response unit output gains were set to $\gamma_P = 6$, which represents an intermediate gain—resulting in neither highly thresholded, nor linear input–output neural function, likely characteristic of average neural populations in the brain (see Fig. 9). Conditioned and reinforcement (*PEn* and *PR*) unit gains were set to a higher value $\gamma_C = 8$, representing a biologically-viable high responsiveness of these units to conditioned input signals. Reinforcement unit input bias was set to $Rb = 0.65$, which results in relatively high (but not maximal) output of the unit during the task. This is considered to represent relatively high dopaminergic stimulation of the PFC to maintain the task-set during performance. Minimal output gain of the task-set units was specified to $\gamma_{Tmin} = 0.5$. This results in almost linear relationship between inputs and outputs for task-set units, with small maximal outputs (e.g. Fig. 9), simulating low PFC activation, and thus low lateral competition when no dopaminergic stimulation is present.

Table 1 Model unit activation parameters

Parameter	Description
Ib	Input unit input bias (fixed)
Tb	Task-set unit input bias (fixed for control simulations)
Rb	Reinforcement prediction unit input bias
Eb	Conditioned processing unit input bias (zero for control simulations)
τ	Time constant of all neural units
γ_P	Output gain of all input, processing and response units
γ_C	Output gain of the conditioned processing units (*PEn* and *PR*)
γ_{Tmin}	Minimal output gain of the task-set units
R_{Th}	Response threshold of the model

Output of each unit is driven by its activation (a) and defined by Eq. 3 (see main text):

$$o_{j,t} = \frac{1}{1 + e^{-\gamma_i\left(Sa_{i,t-1} - \theta\right)}} - d \tag{3}$$

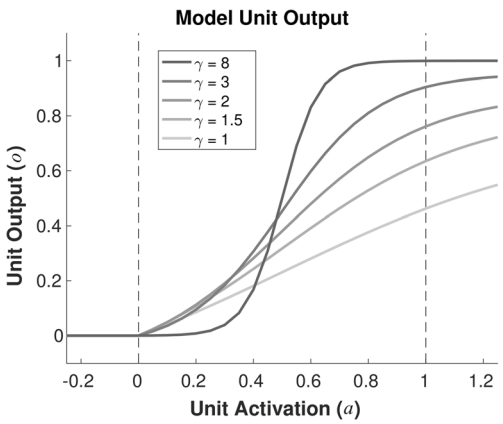

Fig. 9 Model unit output function dependency on unit gain (γ) and unit activation (α)

Table 2 Model connection parameters

Parameter	Description
IPc	Colour input to processing connection strength
PRc	Colour processing to response connection strength
TS	Training scale parameter (*IPw* = *IPc* × *TS* and *PRw* = *PRc* × *TS*)
IPe	Negative word input to conditioned processing connection strength
PTe	Conditioned processing to task-set (negative concept) connection strength
Per	Conditioned processing to reinforcement prediction inhibitory connection strength
r_T	Reinforcement prediction influence scale
Li	Lateral inhibitory connections strength
TcP	Colour-naming task to colour-processing connection strength
TcPw	Colour-naming task to word-processing (*PWr*, *PWg*) connection strength
TwP	Word-reading task to word-processing connection strength
TwR	Word-reading task to response connection strength

In Eq. 3, θ = 1 and *S* = 2 fixed for all units. This constrains unit outputs to lie between zero and one and forces unit output to zero when no unit input is present (see Fig. 9).

Unit output is dependent on the unit gain (γ), which is representative of dopamine levels for task-set (PFC) units (Cohen & Servan-Schreiber, 1993; Servan-Schreiber et al., 1990; Servan-Schreiber et al., 1998). Figure 9 illustrates effect of the γ parameter over the unit output function. As can be noted from Fig. 9, high unit gain results in lower outputs at low activations (e.g. γ = 8, activation below 0.5), but higher maximal output (γ = 8, activation above 0.8). Lower gain, on the other hand, results in higher unit outputs at low activations, but lower outputs at higher activations, with lower maximal output (γ = 2). Lower gain thus increases outputs of less active units and decreases outputs of highly active units. This promotes competition between newly-activated (e.g. *Tn*) and already highly active units (e.g. *Tc*) when they are inhibitorily interconnected. Negative conditioned stimuli in our model lead to decreased gain of the task-set (PFC) units—which we suggest facilitates competition—activation of the negative-concept (*Tn*) unit and deactivation of the current task-set (e.g. *Tc*) unit.

Model connection parameters are briefly described in Table 2. Constraints applied to specify the model connection parameters are described in the main methods section of the report. The specified model connection parameters are presented in Table 3.

Appendix 2 | Depression model

Parameters of units and connections responsible for task-set and conditioned stimulus processing have been considered as relevant for depression and explored to identify the most depression-relevant mechanism.

Input biases: *Tb, Rb, Eb.*
Output gains: τ_E (output gain of the *PEn* unit), τ_R (output gain of the *PR* unit).
Connection strengths: *IPe, PTe, Per, Rt.*

Alteration in no single parameter could account for the entire pattern of depressive behavioural deficits.

Primary identified simple depression mechanism:

1. Increase in conditioned processing unit input bias *Eb* from zero to 0.615 (amygdala hyperactivity).
2. Increase in conditioned processing to reinforcement prediction connection strength *Per* from -0.25 to -0.33 (increase in amygdala inhibitory influence over dopaminergic transmission).

Alternative identified mechanisms with the amygdalar hyperactivity resulting in equivalent depressive behavioural deficits:

1. Increase in *Eb* from zero to 0.63 and increase in reinforcement prediction unit output gain τ_R from 8 to 14.5 (stronger VTA responsiveness to input).
2. Increase in *Eb* from zero to 0.59 and decrease in connection strength from reinforcement prediction to task-set *Rt* from 8 to 2.5 (weak VTA to PFC connectivity).
3. Increase in *Eb* from zero to 0.60 and decrease in reinforcement prediction unit input bias *Rb* from 0.65 to 0.57 (lower baseline VTA activity).

We consider the *primary* identified mechanism most plausible due to its highest relevance to the neurobiology of depression. Experimental evidence favouring the primary identified depression mechanism over the alternatives is briefly overviewed in the Discussion section.

Table 3 Specified model connection parameters

Parameter	*IPc*	*PRc*	*TS*	*IPe*	*PTe*	*Per*	r_T	*Li*	*TcP*	*TcPw*	*TwP*	*TwR*
Value	0.5	0.8	1.2	1.0	1.0	- 0.25	8.0	- 0.8	0.75	0.33	1.1	0.3

References

Abler, B., Erk, S., Herwig, U., & Walter, H. (2007). Anticipation of aversive stimuli activates extended amygdala in unipolar depression. *Journal of Psychiatric Research, 41,* 511–522.

Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General, 133,* 323–338.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association Press.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences, 12,* 193–200.

Banich, M. T., Milham, M. P., Atchley, R., Cohen, N. J., Webb, A., Wszalek, T., . . . Magin R. (2000). fMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *Journal of Cognitive Neuroscience, 18,* 242–257.

Banich, M. T., Milham, M. P., Jacobson, B. L., Webb, A., Wszalek, T., Cohen, N. J., & Kramer, A. F. (2001). Attentional selection and the processing of task-irrelevant information: Insights from fMRI examinations of the Stroop task. *Progress in Brain Research, 134,* 459–470.

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York, NY: Harper & Row.

Bellaiche, L., Asthana, M., Ehlis, A., Polak, T., & Herrmann, M. (2013). The modulation of error processing in the medial frontal cortex by transcranial direct current stimulation. *Neuroscience Journal, 2013,* 187692. doi:10.1155/2013/187692

Blier, P., & El Mansari, M. (2013). Serotonin and beyond: Therapeutics for major depression. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 368*(1615), 20120536. doi:10.1098/rstb.2012.0536

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108,* 624–652.

Bourdy, R., & Barrot, M. (2012). A new control center for dopaminergic systems: Pulling the VTA by the tail. *Trends in Neurosciences, 35,* 681–690.

Braver, T. S., Barch, D. M., & Cohen, J. D. (1999). Cognition and control in schizophrenia: A computational model of dopamine and prefrontal function. *Biological Psychiatry, 46,* 312–328.

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational control: Rewarding, aversive, and alerting. *Neuron, 68,* 815–834.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences, 4,* 215–222.

Bracht T., Linden D., & Keedwell P. (2015). A review of white matter microstructure alterations of pathways of the reward circuit in depression. *Journal of Affective Disorders, 187,* 45–53.

Cabib, S., & Puglisi-Allegra, S. (1996). Stress, depression and the mesolimbic dopamine system. *Psychopharmacology, 128*(4), 331–342.

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience and Biobehavioral Reviews, 26,* 321–352.

Carmichael, S. T., & Price, J. L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal of Comparative Neurology, 363,* 615–641.

Chang, C. H., & Grace, A. A. (2014). Amygdala-ventral pallidum pathway decreases dopamine activity after chronic mild stress in rats. *Biological Psychiatry, 76,* 223–230.

Cohen, J. D., Braver, T. S., & O'Reilly, R. C. (1996). A computational approach to prefrontal cortex, cognitive control and schizophrenia: Recent developments and current challenges. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 351,* 1515–1527.

Cohen, J. D., Dunbar, K., & McClelland, J. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361.

Cohen, J. D., & Servan-Schreiber, D. (1993). A theory of dopamine function and its role in cognitive deficits in schizophrenia. *Schizophrenia Bulletin, 19,* 85–104.

Compton, R. J., Banich, M. T., Mohanty, A., Milham, M. P., Herrington, J., Miller, G. A., . . . Heller W. (2003). Paying attention to emotion: An fMRI investigation of cognitive and emotional Stroop tasks. *Cognitive, Affective, and Behavioural Neuroscience, 3,* 81–96.

Cools, R. (2008). Role of dopamine in the motivational and cognitive control of behavior. *The Neuroscientist, 14,* 381–395.

Cooney, R. E., Joormann, J., Eugène, F., Dennis, E. L., & Gotlib, I. H. (2010). Neural correlates of rumination in depression. *Cognitive, Affective, and Behavioral Neuroscience, 10,* 470–478.

Courtin, J., Bienvenu, T. C., Einarsson, E. Ö., & Herry, C. (2013). Medial prefrontal cortex neuronal circuits in fear behavior. *Neuroscience, 240,* 219–242.

DeRubeis, R. J., Siegle, G. J., & Hollon, S. D. (2008). Cognitive therapy versus medication for depression: Treatment outcomes and neural mechanisms. *Nature Reviews Neuroscience, 9,* 788–796.

Disner, S. G., Beevers, C. G., Haigh, E. A., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience, 12,* 467–477.

Drevets, W. C. (2003). Neuroimaging abnormalities in the amygdala in mood disorders. *Annals of the New York Academy of Sciences, 985,* 420–444.

Dreyer, J. K., Herrik, K. F., Berg, R. W., & Hounsgaard, J. D. (2010). Influence of phasic and tonic dopamine release on receptor activation. *Journal of Neuroscience, 30*(42), 14273–14283.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 622–639.

Dunlop, B. W., & Nemeroff, C. B. (2007). The role of dopamine in the pathophysiology of depression. *Archives of General Psychiatry, 64*(3), 327–337.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34,* 917–928.

Egner, T., Etkin, A., Gale, S., & Hirsch, J. (2008). Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. *Cerebral Cortex, 18,* 1475–1484.

Eldar, E., Cohen, J. D., & Niv, Y. (2013). The effects of neural gain on attention and learning. *Nature Neuroscience, 16*(8), 1146–1153.

Epp, A. M., Dobson, K. S., Dozois, D. J., & Frewen, P. A. (2012). A systematic meta-analysis of the Stroop task in depression. *Clinical Psychology Review, 32,* 316–328.

Etkin, A., Egner, T., Peraza, D., Kandel, E., & Hirsch, J. (2006). Resolving emotional conflict: A role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron, 51,* 871–882.

Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological Review, 113,* 300–326.

Frank, M. J., Cohen, M. X., & Sanfey, A. G. (2009). Multiple systems in decision making: A neurocomputational perspective. *Current Directions in Psychological Science, 18,* 73–77.

Frings, C., Englert, J., Wentura, D., & Bermeitinger, C. (2010). Decomposing the emotional Stroop effect. *Quarterly Journal of Experimental Psychology, 63,* 42–49.

Friston, K. J. (2009). Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLOS Biology, 7*(2), e33.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage, 19*(4), 1273–1302.

Garris, P. A., & Wightman, R. M. (1994). Different kinetics govern dopaminergic transmission in the amygdala, prefrontal cortex, and striatum: An in vivo voltammetric study.Journal of. *Neuroscience, 14*(1), 442–450.

Gessa, G. (1996). Dysthymia and depressive disorders: Dopamine hypothesis. *European Psychiatry, 11,* 123s–127s.

Gotlib, I. H., & McCann, C. D. (1984). Construct accessibility and depression: An examination of cognitive and affective factors. *Journal of Personality and Social Psychology, 47,* 427–439.

Grabenhorst, F., & Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends in Cognitive Sciences, 15,* 56–67.

Hamann, S., & Mao, H. (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *Neuroreport, 13,* 15–19.

Hamilton, J. P., Etkin, A., Furman, D. J., Lemus, M. G., Johnson, R. F., & Gotlib, I. H. (2012). Functional neuroimaging of major depressive disorder: A meta-analysis and new integration of base line activation and neural response data. *American Journal of Psychiatry, 169,* 693–703.

Herd, S. A., Banich, M. T., & O'Reilly, R. C. (2006). Neural mechanisms of cognitive control: An integrative model of Stroop task performance and fMRI data. *Journal of Cognitive Neuroscience, 18,* 22–32.

Holland, P. C., & Gallagher, M. (2004). Amygdala-frontal interactions and reward expectancy. *Current Opinion in Neurobiology, 14,* 148–155.

Holmes, A. J., & Pizzagalli, D. A. (2008). Response conflict and frontocingulate dysfunction in unmedicated participants with major depression. *Neuropsychologia, 46,* 2904–2913.

Iordan, A. D., Dolcos, S., & Dolcos, F. (2013). Neural signatures of the response to emotional distraction: A review of evidence from brain imaging investigations. *Frontiers in Human Neuroscience, 7,* 200.

Isenberg, N., Silbersweig, D., Engelien, A., Emmerich, S., Malavade, K., Beattie, B., . . . Stern E. (1999). Linguistic threat activates the human amygdala. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 10456–10459.

IsHak, W. W., Davis, M., Jeffrey, J., Balayan, K., Pechnick, R. N., Bagot, K., & Rapaport, M. H. (2009). The role of dopaminergic agents in improving quality of life in major depressive disorder. *Current Psychiatry Reports, 11,* 503–508.

Jennings, J. H., Sparta, D. R., Stamatakis, A. M., Ung, R. L., Pleil, K. E., Kash, T. L., & Stuber, G. D. (2013). Distinct extended amygdala circuits for divergent motivational states. *Nature, 496,* 224–228.

Jones, N. P., Siegle, G. J., Muelly, E. R., Haggerty, A., & Ghinassi, F. (2010). Poor performance on cognitive tasks in depression: Doing too much or not enough? *Cognitive, Affective, and Behavioural Neuroscience, 10,* 129–140.

Kayser, A. S., Allen, D. C., Navarro-Cebrian, A., Mitchell, J. M., & Fields, H. L. (2012). Dopamine, corticostriatal connectivity, and intertemporal choice. *Journal of Neuroscience, 32,* 9402–9409.

Kikuchi T., Miller J. M., Schneck N., Oquendo M. A., Mann J. J., Parsey R. V., & Keilp J. G. (2012). Neural responses to incongruency in a blocked-trial Stroop fMRI task in major depressive disorder. *Journal of Affective Disorders, 143,* 241–247.

Lammel, S., Lim, B. K., & Malenka, R. C. (2014). Reward and aversion in a heterogeneous midbrain dopamine system. *Neuropharmacology, 76,* 351–359.

Lapish, C. C., Kroener, S., Durstewitz, D., Lavin, A., & Seamans, J. K. (2007). The ability of the mesocortical dopamine system to operate in distinct temporal modes. *Psychopharmacology, 191*(3), 609–625.

LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology, 23,* 727–738.

Levens, S. M., Muhtadie, L., & Gotlib, I. H. (2009). Rumination and impaired resource allocation in depression. *Journal of Abnormal Psychology, 118,* 757–766.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203.

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience, 14,* 154–162.

Mandell, D., Siegle, G. J., Shutt, L., Feldmiller, J., & Thase, M. E. (2014). Neural substrates of trait ruminations in depression. *Journal of Abnormal Psychology, 123,* 35–48.

Marek, R., Strobel, C., Bredy, T. W., & Sah, P. (2013). The amygdala and medial prefrontal cortex: Partners in the fear circuit. *Journal of Physiology, 591,* 2381–2391.

Maren, S. (2005). Synaptic mechanisms of associative memory in the amygdala. *Neuron, 47,* 783–786.

Maroun, M. (2013). Medial prefrontal cortex: Multiple roles in fear and extinction. *The Neuroscientist, 19,* 370–383.

Matthews, G., & Harley, T. (1996). Connectionist models of emotional distress and attentional bias. *Cognition & Emotion, 10,* 561–600.

Mayberg, H. S. (1997). Limbic-cortical dysregulation: A proposed model of depression. *Journal of Neuropsychiatry and Clinical Neurosciences, 9,* 471–481.

McKenna, F. P., & Sharma, D. (2004). Reversing the emotional Stroop effect reveals that it is not what it seems: The role of fast and slow components. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 382–392.

Mileykovskiy, B., & Morales, M. (2011). Duration of inhibition of ventral tegmental area dopamine neurons encodes a level of conditioned fear. *Journal of Neuroscience, 31,* 7471–7476.

Mitterschiffthaler, M. T., Williams, S. C., Walsh, N. D., Cleare, A. J., Donaldson, C., Scott, J., & Fu, C. H. (2008). Neural basis of the emotional Stroop interference effect in major depression. *Psychological Medicine, 38,* 247–256.

Mohanty, A., Engels, A. S., Herrington, J. D., Heller, W., Ho, M. H., Banich, M. T., . . . Miller, G. A. (2007). Differential engagement of anterior cingulate cortex subdivisions for cognitive and emotional function. *Psychophysiology, 44,* 343–351.

Mulinari, S. (2012). Monoamine theories of depression: Historical impact on biomedical research. *Journal of the History of the Neurosciences, 21*(4), 366–392.

Naccache, L., Gaillard, R., Adam, C., Hasboun, D., Clémenceau, S., Baulac, M., . . . Cohen L. (2005). A direct intracranial record of emotions evoked by subliminal words. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 7713–7717.

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: insights from a meta-analysis of neuroimaging tasks. *Cognitive, Affective, and Behavioural Neuroscience, 7,* 1–17.

Nestler, E. J., & Carlezon, W. A., Jr. (2006). The mesolimbic dopamine reward circuit in depression. *Biological Psychiatry, 59,* 1151–1159.

Oleson, E. B., & Cheer, J. F. (2013). On the role of subsecond dopamine release in conditioned avoidance. *Frontiers in Neuroscience, 7,* 96.

Oleson, E. B., Gentry, R. N., Chioma, V. C., & Cheer, J. F. (2012). Subsecond dopamine release in the nucleus accumbens predicts conditioned punishment and its successful avoidance. *Journal of Neuroscience, 32,* 14804–14808.

Peckham, A. D., McHugh, R. K., & Otto, M. W. (2010). A meta-analysis of the magnitude of biased attention in depression. *Depression and Anxiety, 27,* 1135–1142.

Phaf, R. H., & Kan, K. J. (2007). The automaticity of emotional Stroop: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 38,* 184–199.

Phelps, E. A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology, 57,* 27–53.

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron, 48,* 175–187.

Pizzagalli, D. A. (2014). Depression, stress, and anhedonia: Toward a synthesis and integrated model. *Annual Review of Clinical Psychology, 10,* 393–423.

Raio, C., Carmel, D., Carrasco, M., & Phelps, E. A. (2012). Nonconscious fear is quickly acquired but swiftly forgotten. *Current Biology, 22*(12), R477–R449.

Rampello, L., Nicoletti, F., & Nicoletti, F. (2000). Dopamine and depression: Therapeutic implications. *CNS Drugs, 13*(1), 35–45.

Ray, R. D., & Zald, D. H. (2012). Anatomical insights into the interaction of emotion and cognition in the prefrontal cortex. *Neuroscience and Biobehavioral Reviews, 36,* 479–501.

Roiser, J. P., & Sahakian, B. J. (2013). Hot and cold cognition in depression. *CNS Spectrums, 18,* 139–149.

Seamans, J. K., & Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in Neurobiology, 74*(1), 1–58.

Servan-Schreiber, D., Bruno, R. M., Carter, C. S., & Cohen, J. D. (1998). Dopamine and the mechanisms of cognition: Part I. A neural network model predicting dopamine effects on selective attention. *Biological Psychiatry, 43,* 713–722.

Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A network model of catecholamine effects: Gain, signal-to-noise ratio, and behavior. *Science, 249,* 892–895.

Shultz, E., & Malone, D. A., Jr. (2013). A practical approach to prescribing antidepressants. *Cleveland Clinic Journal of Medicine, 80,* 625–631.

Straube, T., Sauer, A., & Miltner, W. H. (2011). Brain activation during direct and indirect processing of positive and negative words. *Behavioural Brain Research, 222,* 66–72.

Tanaka, K., Furuyashiki, T., Kitaoka, S., Senzai, Y., Imoto, Y., Segi-Nishida, E., . . . Narumiya S. (2012). Prostaglandin E2-mediated attenuation of mesocortical dopaminergic pathway is critical for susceptibility to repeated social defeat stress in mice. *Journal of Neuroscience, 32*(12), 4319–4329.

Thurley, K., Senn, W., & Lüscher, H. R. (2008). Dopamine increases the gain of the input-output response of rat prefrontal pyramidal neurons. *Journal of Neurophysiology, 99,* 2985–2997.

Tye, K. M., Mirzabekov, J. J., Warden, M. R., Ferenczi, E. A., Tsai, H. C., Finkelstein, J., . . . Deisseroth K. (2013). Dopamine neurons modulate neural encoding and expression of depression-related behaviour. *Nature, 493,* 537–541.

Volman, S. F., Lammel, S., Margolis, E. B., Kim, Y., Richard, J. M., Roitman, M. F., & Lobo, M. K. (2013). New insights into the specificity and plasticity of reward and aversion encoding in the mesolimbic system. *Journal of Neuroscience, 33,* 17569–17576.

Wagner G., Sinsel E., Sobanski T., Köhler S., Marinou V., Mentzel H. J., Sauer H., & Schlösser R. G. (2006). Cortical inefficiency in patients with unipolar depression: an event-related fMRI study with the Stroop task. *Biological Psychiatry, 59,* 958–965.

Watkins, E., & Brown, R. G. (2002). Rumination and executive function in depression: An experimental study. *Journal of Neurology, Neurosurgery, and Psychiatry, 72,* 400–402.

Wayment, H. K., Schenk, J. O., & Sorg, B. A. (2001). Characterization of extracellular dopamine clearance in the medial prefrontal cortex: Role of monoamine uptake and monoamine oxidase inhibition. *Journal of Neuroscience, 21*(1), 35–44.

Whalen, P. J., Bush, G., McNally, R. J., Wilhelm, S., McInerney, S. C., Jenike, M. A., & Rauch, S. L. (1998). The emotional counting Stroop paradigm: A functional magnetic resonance imaging probe of the anterior cingulate affective division. *Biological Psychiatry, 44,* 1219–1228.

Whalen, P. J., Shin, L. M., Somerville, L. H., McLean, A. A., & Kim, H. (2002). Functional neuroimaging studies of the amygdala in depression. *Seminars in Clinical Neuropsychiatry, 7,* 234–242.

Williams, J. M., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin, 120,* 3–24.

Wyble, B., Sharma, D., & Bowman, H. (2008). Strategic regulation of cognitive control by emotional salience: A neural network model. *Cognition & Emotion, 22,* 1019–1051.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111,* 931–959.